Investigation of the translation-initiation factor IF2 gene, *infB*, as a tool to study the population structure of *Streptococcus agalactiae*

Jakob Hedegaard,¹ Majbritt Hauge,² Jeppe Fage-Larsen,¹
Kim Kusk Mortensen,¹ Mogens Kilian,² Hans Uffe Sperling-Petersen¹
and Knud Poulsen²

Author for correspondence: Knud Poulsen. Tel: +45 89421736. Fax: +45 86196128.
e-mail: kp@microbiology.au.dk

¹ Department of
Biostructural Chemistry,
Institute of Molecular and
Structural Biology, Aarhus
University, Gustav
Wiedsvej 10C, DK-8000
Aarhus C, Denmark

² Department of Medical
Microbiology and
Immunology, The
Bartholin Building,
Aarhus University, DK-8000
Aarhus C, Denmark

The sequence of *infB*, encoding the prokaryotic translation-initiation factor 2 (IF2), was determined in eight strains of *Streptococcus agalactiae* (group B streptococcus) and an alignment revealed limited intraspecies diversity within *S. agalactiae*. The amino acid sequence of IF2 from *S. agalactiae* and from related species were aligned and revealed an interspecies conserved central and C-terminal part, and an N-terminal part that is highly variable in length and amino acid sequence. The diversity and relationships in a collection of 58 genetically distinct strains of *S. agalactiae* were evaluated by comparing a partial sequence of *infB*. A total of six alleles were detected for the region of *infB* analysed. The alleles correlated with the separation of the same strains of *S. agalactiae* into major evolutionary lineages, as shown in previous work. The partial sequences of *infB* were furthermore used in phylogenetic analyses of species closely related to *S. agalactiae*, yielding an evolutionary tree which had a topology similar to a tree constructed using 16S rRNA sequences from the same species.

Keywords: *Streptococcus agalactiae*, *infB*, initiation factor 2, clonal population structure

INTRODUCTION

In the past 25 years *Streptococcus agalactiae*, also known as group B streptococci, has emerged as the leading cause of illness and death among newborn infants (Schuchat, 1998). An estimated 7500 episodes of neonatal sepsis and meningitis occurred in 1998 in the USA (information provided by the Centers for Disease Control and Prevention, the National Center for Infectious Diseases and the Division of Bacterial and Mycotic Diseases). *S. agalactiae* is part of the normal microflora in the vagina in many women from where it may colonize the newborn during delivery. It is not completely understood why some infants develop disease upon colonization. At present, no vaccine preventing invasive diseases caused by *S. agalactiae* is

available and intrapartum prophylactic treatment with antibiotics is used routinely in some countries (Schuchat, 1999). The polysaccharide capsule is recognized as a major virulence factor in *S. agalactiae* and, in addition, the bacterium may express a number of proteins including C5a peptidase, hyaluronidase, alpha-antigen or RIP, and beta-antigen that are potential virulence factors (Cleat & Timmis, 1987; Li *et al.*, 1997; Pritchard & Lin, 1993; Stålhammar-Carlemalm *et al.*, 1993; Wexler *et al.*, 1985). Differences in pathogenic potential among subpopulations of *S. agalactiae* have been disputed. Musser *et al.* (1989) identified a high-virulence clonal type of *S. agalactiae* serotype III in a collection of strains from the USA, Quentin *et al.* (1995) found that among French isolates three clonal families were responsible for the majority of invasive disease in neonates and Granlund *et al.* (1998) described an association between strains characterized by the presence of an insertion element within the hyaluronidase gene and endocarditis. We found that a population of *S. agalactiae*

The EMBL accession numbers for the sequences reported in this paper are AJ003164 and AJ251493 to AJ251499.

strains isolated in Denmark was predominantly clonal and there was no evidence of association between specific evolutionary lines and neonatal disease (Hauge *et al.*, 1996). Additional molecular markers with a suitable variation within the species would be valuable tools for epidemiological and evolutionary studies on *S. agalactiae* to evaluate the discrepancies described above. Sequence variation in *infB*, encoding the prokaryotic translation-initiation factor 2 (IF2), has been studied in different isolates of *Escherichia coli* (Steffensen *et al.*, 1997) and the part of *infB* encoding the GTP-binding domain of IF2 has been found to contain adequate variation for use as a phylogenetic marker within the family *Enterobacteriaceae* (Hedegaard *et al.*, 1999).

IF2 interacts with fMet-tRNA^{Met}_f, GTP, IF1, IF3 and both 30S and 50S ribosomal subunits. Through these interactions, IF2 promotes the binding of the initiator tRNA to the 30S ribosomal subunit and catalyses the hydrolysis of GTP following 70S initiation-complex formation. Initiation-complex formation has been extensively reviewed (Gold, 1988; Gualerzi & Pon, 1990; Hershey, 1987; Maitra *et al.*, 1982). IF2 from *E. coli* exists in three forms: IF2-1 (97.3 kDa), IF2-2 (79.7 kDa) and IF2-3 (78.8 kDa). A new nomenclature for translation factors has been proposed by International Union of Biochemistry and Molecular Biology (1996). According to this new nomenclature, IF2 α , IF2 β and IF2 γ are named IF2-1, IF2-2 and IF2-3, respectively. The IF2-2 and IF2-3 forms are both translated from in-frame translation-initiation codons in the single-copy gene *infB*. Outside the family *Enterobacteriaceae*, internal translation initiation within *infB* has been found in *Bacillus subtilis*, where two forms of IF2 were detected: IF2-1 (78.6 kDa) and IF2-2 (68.2 kDa) (Hubert *et al.*, 1992; Shazand *et al.*, 1990).

In this study, we confirm the separation of 58 strains of *S. agalactiae* into separate evolutionary lineages as previously reported by Hauge *et al.* (1996). Furthermore, we found that a partial sequence of *infB* was useful as a molecular marker for phylogenetic analysis of *S. agalactiae* and closely related species. The amino acid sequence deduced from the full-length sequence of *infB* from eight strains of *S. agalactiae* verified the high interspecies variation in the N-terminal sequence of IF2 and the conserved C-terminal region. The predicted start codon of *infB* in *S. agalactiae* was verified by subcloning and overexpression in *E. coli* followed by purification and N-terminal sequencing of an N-terminal fragment of IF2 from *S. agalactiae*.

METHODS

Bacterial strains and cloning vectors. The *S. agalactiae* strains used in this study for complete sequencing of the *infB* gene were 3164, 941, 3114, 3076, 949, 3074, 3163 and 3098 of serotypes III, Ia, Ib, II, III, III and Ia, respectively, representing the clonal types 4, 13, 15, 25, 37, 43, 52 and 58, respectively, described by Hauge *et al.* (1996). Strains 3164 and 3163 represented the so-called low and high virulent types, respectively, isolated in the USA (Musser *et al.*, 1989), whereas

the remaining strains were human isolates from Denmark. In addition, 50 strains representing each of the remaining clonal types described by Hauge *et al.* (1996) were included for partial sequencing of the *infB* gene. The strains were grown at 37 °C in atmospheric air plus 5% CO₂ on blood agar or in Todd-Hewitt broth. Bacteriophage λ L47.1 (Loenen & Brammar, 1980) was used as a *Bam*HI substitution vector and recombinant phages were plated on *E. coli* K802 (Wood, 1966) as described by Sambrook *et al.* (1989). *E. coli* UT5600 (Elish *et al.*, 1988) was used as host for the thermo-inducible runaway-replication plasmid pCP40 in combination with the regulatory plasmid pCl857 (Remaut *et al.*, 1983). The expression of genes inserted in pCP40 is under control of the λ p_L promoter and the temperature-sensitive cI repressor. The gene inserted must contain its own translational regulatory elements.

Southern blot analysis. The Southern blots were prepared in a previous study on the population structure of *S. agalactiae* (Hauge *et al.*, 1996) and reused for hybridization at low stringency including a stepwise decline in temperature as described by Poulsen *et al.* (1996).

Construction and screening of the *S. agalactiae* 3076 genomic library. Whole-cell DNA from *S. agalactiae* strain 3076 was prepared as described by Hauge *et al.* (1996) and a partial *Sau*3AI digest was fractionated by agarose gel electrophoresis. Fragments in the size range from 10 to 20 kb were extracted from the gel by electroelution and used to prepare a genomic library using λ L47.1 as a *Bam*HI substitution vector (Sambrook *et al.*, 1989). The recombinant phages were packaged *in vitro* with Gigapack II Packaging Extracts (Stratagene) and plated on *E. coli* K802. Positive plaques were identified by *in situ* hybridization using a DNA fragment labelled with [³²P]dATP as probe. The low-stringency conditions used in the hybridization were as described previously (Poulsen *et al.*, 1996). A single positive plaque was purified by replating on *E. coli* K802 and phage DNA was isolated from a 20 ml phage lysate (Mikkelsen *et al.*, 1985).

DNA sequencing. λ phage DNA and different segments of the *infB* gene amplified by PCR served as template in the sequencing reactions. The PCRs were done using Thermo-Prime Plus DNA Polymerase (Advanced Biotechnologies) or ReadyToGo PCR beads (Amersham Pharmacia Biotech) following the protocols supplied with the enzymes. Degenerate oligonucleotide primers for PCR and sequencing were designed by back-translation of selected conserved stretches in an alignment of IF2 sequences from species closely related to *S. agalactiae*. Specific primers for PCR and sequencing were made as the sequence of *infB* from *S. agalactiae* was obtained. Primers for amplifying and sequencing the central variable part of *infB* were 5'-TACTGAGGGCATGACCGTTGC-3' and 5'-GACACCCGACGCTTTAGAGTGAT-3'. All oligonucleotide primers were purchased from DNA Technology. The amplicons were purified for sequencing as described by Hedegaard *et al.* (1999) or using the Wizard Minicolumns purification system (Promega). The sequencing was performed using the Thermo Sequenase fluorescent labelled primer cycle sequencing kit (Amersham Pharmacia Biotech) and analysed on an ALF DNA sequencer (Amersham Pharmacia Biotech) or using Thermo Sequenase dye terminator cycle sequencing kit (Amersham Pharmacia Biotech) and analysed on an ABI 377 DNA sequencer (Perkin Elmer).

Sequence analyses. Sequences obtained were analysed and edited using programs contained in the GCG software package (Genetics Computer Group, University of Wisconsin, Madi-

son, WI, USA). Phylogenetic trees were constructed using the distance-matrix method and parsimony analysis. The distances were corrected using the Kimura two-parameter method (Kimura, 1980). All trees were generated by heuristic searches with tree-bisection branch-swapping and resampled with 100 bootstrap replications to test the robustness of the data. A tree based on 16S rRNA gene sequences was constructed and compared to the one obtained from *infB* sequence data. All phylogenetic analyses were performed with PAUP 4.0.0d55 for UNIX (Distributed by D. L. Swofford & S. Olson, Laboratory of Molecular Systematics, Smithsonian Institution).

Expression in *E. coli*, purification and characterization of an N-terminal fragment of IF2 from *S. agalactiae*. An 1108 bp fragment containing the translational-regulatory region and the part of *infB* encoding the first 330 aa of IF2 from *S. agalactiae* strain 3076 was termed *S. agalactiae* *infB*Δ991–2784. This fragment was ligated into pCP40 as an *EcoRI*/*Bam*HI-digested PCR product made using the primers *EcoRI*-Forward (5'-GGCAGGGAATTCAAGAAAAGTGGTTGCTGTC-GC-3') and *Bam*HI-Reverse (5'-GGGACGTGGATCCTGT-TACTGGTTATGGAG-3'). The obtained plasmid, pCP40(*S. agalactiae* *infB*Δ991–2784), was transformed into competent *E. coli* UT5600(pcl857) cells, prepared as described by Chung *et al.* (1989); plated on 2 × TY agar containing 100 µg ampicillin ml⁻¹ and 50 µg kanamycin ml⁻¹ and incubated overnight at 30 °C. Colonies were screened for the correct insert by PCR using the primers pCP40-Forward (5'-ACTG-GCGGTGATACTGAGC-3') and Saga819-Reverse (5'-TT-TACGACGACTATCTGCTGT-3'). An overexpression test was performed by inoculating 3 ml 2 × TY containing 100 µg ampicillin ml⁻¹ and 50 µg kanamycin ml⁻¹ with selected colonies and incubating at 30 °C for 7 h, followed by 42 °C for 1 h and analysing for overexpression by SDS-PAGE. DNA sequencing of plasmid DNA using the pCP40-Forward primer furthermore verified the presence of the correct insert.

Colonies of *E. coli* UT5600[pcl857, pCP40(*S. agalactiae* *infB*Δ991–2784)] were inoculated into 2 × TY containing 100 µg ampicillin ml⁻¹ and 50 µg kanamycin ml⁻¹, and grown overnight at 30 °C. The overnight culture was diluted 1:100 into 2 × TY containing 100 µg ampicillin ml⁻¹ and 50 µg kanamycin ml⁻¹ and grown at 30 °C to an OD₅₅₀ of 1. Overexpression of the N-terminal fragment, termed *S. agalactiae* IF2Δ331–927, was induced by diluting the cultures with an equal volume of 2 × TY medium containing 100 µg ampicillin ml⁻¹ and 50 µg kanamycin ml⁻¹ at 56 °C, followed by growth for 1.5 h at 42 °C. The cells were harvested by centrifugation at 2200 *g* for 20 min at 4 °C and the cell pellet was washed with 0.9% NaCl. Samples were analysed for overexpression by SDS-PAGE. The harvested cells were resuspended in buffer H(300) [Buffer H is 50 mM HEPES, pH 7.6, 1 mM DTT, 0.1 mM PMSE, 15 mM NaN₃ and NaCl as indicated in parentheses, e.g. H(300) has an NaCl concn of 300 mM] and disrupted by passing through an Aminco French pressure cell at 1500 p.s.i. (10350 kPa). The cell extract was clarified by 1 h centrifugation at 30000 *g* at 4 °C. *S. agalactiae* IF2Δ331–927 was captured on an S-Sepharose Fast Flow 50/3 column (Amersham Pharmacia Biotech), moderately purified on an S-Sepharose High Performance 16/10 column (Amersham Pharmacia Biotech), concentrated in Centricon30 (Amicon), highly purified in an AcA54 16/92 column (Biosepra) and finally concentrated in Centricon30. All chromatographic steps were performed at 4 °C using an ÄKTAexplorer system (Amersham Pharmacia Biotech). The S-Sepharose FF and S-Sepharose HP columns were equilibrated with H(300) before loading the sample. The FF column was used at a flow rate of

7 ml min⁻¹ and bound protein was eluted in a linear gradient from H(300) to H(800) in 3.5 column vols. The HP column was used at a flow rate of 2 ml min⁻¹ and bound protein was eluted in a step gradient from H(300) to H(430) in 4 column vols, to H(530) in 0.3 column vols and to H(1000) in 4 column vols. Isocratic elution with H(100) from the AcA54 column was performed at 0.2 ml min⁻¹. During the purification, absorption at 280, 270 and 254 nm were measured on-line and fractions were analysed by SDS-PAGE. The final protein concentration was determined by measuring A₂₈₀. Purified *S. agalactiae* IF2Δ331–927 was sequenced at the N terminus by automated Edman degradation essentially as described by Nyengaard *et al.* (1991).

RESULTS AND DISCUSSION

Sequencing and characterization of *infB* from *S. agalactiae* strain 3076

We have previously cloned and characterized the *infB* gene encoding IF2 from *Lactococcus lactis* (accession no. AJ005118). In the present study we have used a DNA probe encoding the GTP-binding domain of *L. lactis*

Fig. 1. Coomassie brilliant blue R250 stained SDS-PAGE summarizing the purification of *S. agalactiae* IF2Δ(331–927). MW: molecular mass marker (Amersham Pharmacia Biotech); S30: sample from supernatant after centrifugation of cell extract at 30000 *g*; S-FF: sample of pooled fractions after chromatography on S-Sepharose-FF; S-HP: sample of pooled and concentrated fractions after chromatography on S-Sepharose-HP; AcA: sample of pooled and concentrated fractions after chromatography on AcA54. The fragment has a calculated molecular mass of 32.9 kDa but migrates as an ~40 kDa protein due to the very basic pI of the fragment.

<i>B. subt</i>	1	MAKRVVPEAKALNVSSKILTLALKNDLVANHMAMLEKAIKKLDAKY.....KKGAAAK.....SQKPA...
<i>E. faec</i>	1	MSNKKRIYELAKEINQPSKQVWEKAOQLGIVNKHMTITTCDEKKLQAN.....KKPQTNKKKPAQASQKPAATNQ
<i>S. agal</i>	1	MSKKRLHEIAKEIKTSKSEVVEAQSILGLPVKSHASSVEENATRIVESFS.SSKTKAPTNSVQTNQGVKTESKIVETQGLSSDDKESSTOPVAKPKPQS
<i>S. pyog</i>	1	MSKKRLHEIAKEIKTSKSEVVEHAKYLGLDVKSHASSVEEADAKKIISSESKASKPDVTASQTVKPKVEAQPSTVYVKTGE.SEHVEKTO.VSKPK..S
<i>L. lact</i>	1	MSDKKRINQIAKETGLSTELVATAQSILGFVKSHSSSVTAEEAKSKIIQGVKTGTDITIVNLLKKLLKLLRLVPEAKSKQ...EDHRTI.FAGVAIVED
<i>B. subt</i>	63	ET.....NK.....NKOPQG.....V.....
<i>E. faec</i>	72	PNBQKKOBTKTQCKNNR.....NKOPQG.....V.....
<i>S. agal</i>	99	RNFKAERARAKAEAEKROHNGD.....HRKNNRHNDTRSDRR..HQQQ.....KRSN.GNRND.NRO..GQON.....NRNKNDGRYAD
<i>S. pyog</i>	96	RNFKAERARAKEQAARKQANGSSHSRQERRGGYRQPNNHQTNEQGDKRITHRSGQDNDKRIERKASNVSPRHD.NHOLVDORNRSFAKENHKNGRFTN
<i>L. lact</i>	97	PAIL...ARIKEKEEAQKAAKTEAPEIEVITTEKPKVAEPVKKSEPKAAAKAE.ETKVEKVEAKANTVTPKAEVKTENVADKKEPVVTEKEKKSLTQK
<i>B. subt</i>	74	NQ...QS.....AGNQPNK.....
<i>E. faec</i>	101	NQKNOSTNQNRNNGGNNQNRQCATQSGN.....
<i>S. agal</i>	169	HQKQKPTQRPQAPGNRIDFKARAAALKAENAEYSRHSERQREDEAKRAQAAKE...QELAKAAALKADEAQKAKEKIASKPVAKVKEIVNKVVAATPS
<i>S. pyog</i>	195	QKKQGRGE PQSKS.PKIDFKARAAALKAENAEYSRQSETRFPAQAEAKRLAELA...RQEAKEAALKAQAEEMSHRE.AALKSIEEAETKLKSSNISAK
<i>L. lact</i>	192	PRIQIKVIKRAEDIKKEQAARPEKKFEDKNRNDNRNNDNRPNONGNGQHSQGGNHYNDRPAGQGNQCK.RDKFASGSSSTSDTFTPAASGKN
<i>B. subt</i>	85	IFDGKKNQVQNNQFNKNGKNNNNKKNKNNNN.....NKNOHQQKP.VKP...KK..ELPEKITFSGLTIVGALAEELG
<i>E. faec</i>	132	NRQGSTGGNQNRQGSNQGNNQNRNNNNNRGKNNNNNN.RNRKKGKKGQQTSNKPAVEP...RKRPFLPVELEYTEGMNVADIAKRIH
<i>S. agal</i>	266	QTADSRKKQTRSDKSRQFSNENEDGQKQTKNGKNNNNQVNRNORNSNNNNKKNKKGK...TNGAPKPVTERKPHLPKEFEYTEGMVAELAKRIK
<i>S. pyog</i>	290	STADNRKKQARPEKNRELTHHSQEGCK..KNGKSWNSONQVRNOKNSNNNNKKNKKGKNNVKNNTAPKPVTERKPHLPKEFEYTEGMVAELAKRIK
<i>L. lact</i>	291	NRRDRDRKK...TDSNRD...NTKDCNRKGGPLRVNDNRNOVNARNNSN...NNGGRGR.YQNNQSSNVEATORKFHELPSLEYEVGMVQDIASIK
<i>B. subt</i>	154	KEPSELIKKIMLLGMVMTINQBLDKDTIELIASEYGVTEEVIVLEETELEYEPPD...NEEDLEIRPFVVIMGHVDHGKTTLLDSIRKTKVVEGEAG
<i>E. faec</i>	220	REPABIIKKLFMLGMVNVNQALDKDTIELLATDYGMPQERIQVDIADIDKEFFEBE..LNPDILVSRPPVVTIMGHVDHGKTTLLDTRLHSRVTSGEAG
<i>S. agal</i>	362	REPABIIKKLFMLGMVMTINQSLDCTIELLMVDYGEBAKAVEVDADIERFFADEDYLNPDNLTERPFPVVTIMGHVDHGKTTLLDTRLHSRVATGEAG
<i>S. pyog</i>	388	REPABIIKKLFMLGMVMTINQSLDCTIELLMVDYGEBAKAVEVDADIERFFEDENYLNPNENIVERAPVVTIMGHVDHGKTTLLDTRLHSRVATGEAG
<i>L. lact</i>	382	REPABIIKKLFMLGMVMTINQSLDCTIELLMVDYGVTPKLVPEDKSDIERLVEDGYLNDKVERPAVVTIMGHVDHGKTTLLDRFSRVTEGEAG
<i>B. subt</i>	251	GITQIHIGAYQIEENGKKITFLDTPGHAFTSMRARGASVTDITILVVAADDGVMPQTEVAINHAKAAEVPIIVANKIDKESANDRVMQELTEYGLVPE
<i>E. faec</i>	319	GITQIHIGAYQIDIDCKEITFLDTPGHAFTSMRARGASVTDITILVVAADDGVMPQTEVAINHAKAAAGVPIIVANKIDKPGANPOHVMQELSEYELIPE
<i>S. agal</i>	462	GITQIHIGAYQIEENGKKITFLDTPGHAFTSMRARGASVTDITILVVAADDGVMPQTEVAINHAKAAAGVPIIVANKIDKPGANPERVISELAEEHGHST
<i>S. pyog</i>	488	GITQIHIGAYQIEENGKKITFLDTPGHAFTSMRARGASVTDITILVVAADDGVMPQTEVAINHAKAAAGVPIIVANKIDKPGANPERVIAEAEYGHST
<i>L. lact</i>	482	GITQIHIGAYQIKTNKKITFLDTPGHAFTSMRARGASVTDITILVVAADDGVMPQTEVAINHAKAAAGVPIIVANKIDKPGANPERVTELEHGVFEV
<i>B. subt</i>	351	AWGGE...TIFVPLSALTCKCIDELVEMILLVSEVEELKANFNROAKGTVIEAEALDKGKGSVATLLVQGLTLHVGDPIVVGNTFGRVRAMVNDLGRVRVTA
<i>E. faec</i>	419	AWGGE...TIFVEISAKFGQNIIDELLEMLLVAEVEDLKADPKQRAIGTVIEARLDKKGKGEVTTLLVQOGLTLVQDPIVVGNTYGRVRAMVNDLGRVRKSA
<i>S. agal</i>	562	AWGGE...SEFVEISAKFGKNIIDELLETVLLVAEVEELKADADVRAIGTVIEARLDKKGKAVATLLVQOGLTLVQDPIVVGNTFGRVRAMVNDLGRVRKVA
<i>S. pyog</i>	588	AWGGE...SEFVEISAKFNKNIIDELLETVLLVAEVEELKADDPVRAIGTVIEARLDKKGKATATLLVQOGLTLHVQDPIVVGNTFGRVRAMVNDLGRVRKSA
<i>L. lact</i>	582	AWDPENGSEFVEISAKFNQLEDLDTVLLVAEVEELKADPSVRAIGTVIEARLDKKGKATATLLVQOGLTLHVDPIVVGNTYGRVRAMVNDLGRVRKSA
<i>B. subt</i>	449	GPSTFVEITGLNDVPOAGDQELVFDKERTARSVGEARASKQEEQRSDAKATSLDDLFEEQIKOGDVKDINLTVKADVQGSAAELAAALQKIEVEGVKVKI
<i>E. faec</i>	517	GPATPVEITGLNDVPOAGDRFVVFDEKRTARAAGEERAKRALLHRAITSRVILDNLFESLKEGELKEVNVIIKADVQGSAAELAAASLOKIDVEGVKVKI
<i>S. agal</i>	607	GPSTFVEITGLNEAPVAGDHFVAYADEKAARAAGEERAKRALLKORONTORVSLNLFDTLKAGEVKSNNVIIKADVQGSAAELAAASLIKIDVEGVKVVN
<i>S. pyog</i>	686	EPSTFVEITGLNETP VAGDHFVAYADEKAARAAGEERSKRALLKORONTORVSLNLFDTLKAGEIKTVNVIKADVQGSAAELAAASLVKIEVEGVKVVN
<i>L. lact</i>	682	GPSTFVEITGLSDVPOAGDHFVAFVDEKAARAAGEERAKRAOLIKRONTRRVNLDNLFDTLKAGEOTKSVNIIKADVQGSAAELAAASLOKIEVEGVKVDI
<i>B. subt</i>	549	IITGVGAINESDITILASASNAIIGFNVREDGNAKSTAEAEVDIRLHRIIYKVIDEIEAAMKGMLDPEYEEKVIGQVEVROTFKVSKI GTIAGGYVTEG
<i>E. faec</i>	617	VHSAVGAINESDVTILASASNAIIGFNVRTPOAKIQADNEVDIRLHRIIYKALIEIEIEMKGMLDPEYEEKITQCMIVRETFKVSQVGTIAGGYVTEG
<i>S. agal</i>	760	VHSAVGAINESDVTILASASNAIIGFNVRTPOARQADADVDVIRCHSIIYKVIEEVEBAMKCKLDPEYQEKILGEAIIRETFFKVSQVGTIAGGMVING
<i>S. pyog</i>	786	VHSAVGAINESDVTILASASNAIIGFNVRTPOARQADADVDVIRLHRIIYKVIEEVEBAMKCKLDPEYQEKILGEAIIRETFFKVSQVGTIAGGMVING
<i>L. lact</i>	782	VHSAVGAINESDITILASASNAIIGFNVRTGLAREQAQAEVDIRLHRSIIYKVIEEVEBAMHGMLDPEYEEKIIGEAIVRETFKVSQVGTIAGGMVING
<i>B. subt</i>	649	TITRDSGLRIIRDGVVIFECEVDVILKRFKDDVKEVSQGYSCGITHKKYNDIREGIIIEAFVMOETERT.
<i>E. faec</i>	717	YIRRDSVVRVIRIDGVVIFECEVLAASLKRFKDDVKEVKMGYSCGAMIEKENDIKVDDVIEGHIMEEIKTN.
<i>S. agal</i>	860	KVTRDSSVRVIRIDGVVIFDCKLASLKHYKDDVKEVGNAQEGGLMIENYNDLKEDDTIEAYIMEEIKRK.
<i>S. pyog</i>	886	KVTRDSSVRVIRIDSVVIEDCKLASLKHYKDDVKEVGNAQEGGLMIENYNDLVODTIEAYIMEEIVRK.
<i>L. lact</i>	882	KVTRDASVRVIRIDGVVIFDCAIASLKHFDDVKEVGNAQEGGLMIEDFNDVIDDTIEVVKVIEIRPKK

Fig. 2. Alignment of selected sequences of IF2. *B. subt*, *Bacillus subtilis* (accession no. Z18631; Shazand *et al.*, 1990); *E. faec*, *Enterococcus faecium* (accession no. M36878; Friedrich *et al.*, 1988); *S. agal*, *Streptococcus agalactiae* 3076 (accession no. AJ003164; this work); *S. pyog*, *Streptococcus pyogenes* (The Streptococcal Genome Sequencing Project); and *L. lact*, *Lactococcus lactis* (accession no. AJ005118). The proposed domain borders are shown by >|< and the five essential regions, G1 to G5, of the GTP-binding motif are indicated by a ~. The polymorphic amino acid positions in *S. agalactiae* IF2 are indicated by asterisks.

IF2, i.e. Ile³⁹⁶ to Val⁵⁸⁴, to screen for an *infB* homologue in *S. agalactiae*. Southern blot experiments using low stringency conditions for hybridization showed that

probe recognized a single *EcoRI* restriction fragment of the same size in the genome of all 58 strains analysed (data not shown). Thus, the probe appeared to be highly

Fig. 3. Dendrograms based on sequence data from *infB* and 16S rRNA using the distance method. For both *infB* and 16S rRNA sequences, phylogenetic analyses using the parsimony method resulted in one most parsimonious tree, having a topology identical to the one obtained by the distance method. The numbers at branch sites are the percentages of bootstrap replications supporting the branch. The *infB* tree is based on data from *S. agalactiae*, accession no. AJ003164 (this work); *S. pyogenes* (the streptococcal genome sequencing group); *L. lactis*, accession no. AJ005118; *Ent. faecium*, accession no. M36878 (Friedrich et al., 1988); *B. subtilis*, accession no. Z18631 (Shazand et al., 1990); *B. stearothermophilus*, accession no. X04399 (Brombach et al., 1986); *M. genitalium*, accession no. U39695 (Fraser et al., 1995); *M. pneumoniae*, accession no. AE000062 (Himmelreich et al., 1996) and *Esc. coli*, accession no. K01175 (Sacerdot et al., 1984). The central part of *infB* (bp 1285–1794 in *S. agalactiae* *infB*) was used for phylogenetic analyses. The parsimony analyses showed that of 519 characters, 259 were parsimony informative (202 were identical and 58 were parsimony uninformative). The 16S rRNA tree is based on data from *S. agalactiae*, accession no. AB002479; *S. pyogenes*, accession no. AB002521; *L. lactis*, accession no. M58837; *Ent. faecium*, accession no. Y18294; *B. subtilis*, accession no. AF008220 (Green et al., 1985); *B. stearothermophilus*, accession no. X60640 (Ash et al., 1991); *M. genitalium*, accession no. X77334; *M. pneumoniae*, accession no. M29061 (Weisburg et al., 1989) and *Esc. coli*, accession no. J01695 (Brosius et al., 1978). A segment of the 16S rRNA gene corresponding to nucleotide positions 43–1374 with reference to the *S. agalactiae* 16S rRNA gene was used for the analyses. Positions containing ambiguously determined nucleotides were deleted in all 16S rRNA sequences. The parsimony analyses showed that of 1343 characters, 416 were parsimony informative (760 were identical and 167 were parsimony uninformative).

specific under the conditions used. Subsequently, the same probe and conditions were used to identify a positive clone, λ Sa-IF2-4.4, in a genomic lambda phage library of *S. agalactiae* strain 3076. Recognition sites for the restriction enzymes *Eco*RI and *Hind*III were mapped within the DNA of λ Sa-IF2-4.4. The results showed that this recombinant phage contained an insert of 12.5 kb including an *Eco*RI fragment of 10.6 kb with homology

to the *infB* probe from *L. lactis* as revealed by Southern blot analyses.

A fragment of *infB* encoding the central and conserved part of IF2 was amplified by PCR using degenerate oligonucleotide primers and DNA of phage λ Sa-IF2-4.4 as the template. The amplicons were purified and subsequently sequenced. The terminal parts of *infB* and

Fig. 4. Polymorphic positions detected for *infB* (top) and IF2 (bottom) in eight strains of *S. agalactiae*. Domain borders, as defined by homology to *E. coli* IF2, and functional regions are indicated. The borders between domains I, II and III are dashed because their exact position can not be determined due to lack of homology to *E. coli* IF2. The four allele groups (A–D) are indicated.

flanking regions were sequenced using phage DNA as the template. A sequence of 3391 bp was obtained covering, in addition to *infB*, the gene downstream, *rbfA*, encoding ribosome-binding factor A. The identity of both genes was verified by comparison with known sequences of *infB* and *rbfA* from other species. In *S. agalactiae* strain 3076, *infB* was found to constitute an ORF of 2784 bp starting with a UUG start codon and with the potential of encoding a 102.4 kDa protein of 927 aa with a calculated pI of 9.0. Due to the strongly basic N-terminal region this pI value is exceptionally high, compared to IF2 in other prokaryotes. The proposed translation-start codon was confirmed by heterologous overexpression in *E. coli*, purification and N-terminal sequencing of an N-terminal fragment of *S. agalactiae* IF2. A 1108 bp fragment including sequences upstream of the start codon and encoding the first 330 aa (32.9 kDa, pI 9.9), was subcloned into pCP40 and the obtained plasmid, pCP40(*S. agalactiae* *infB*Δ991–2784), was transformed into competent *E. coli* UT5600(pcl857) followed by overexpression and purification of *S. agalactiae* IF2Δ331–927 (Fig. 1). The first 5 aa of the purified recombinant protein were determined as Ser-Lys-Lys-Arg-Leu, which is in agreement with the deduced amino acid sequence, assuming that the N-terminal fMet is removed by methionyl-aminopeptidase (Hirel *et al.*, 1989).

Homology between IF2 sequences from *S. agalactiae* and related species

IF2 sequences from several bacterial species are available from public databases. A comparison of the sequences shows that this initiation factor varies considerably in size among different species from 571 aa (63.2 kDa) in *Thermus thermophilus* (accession no. Z48001) to 1054 aa (111.3 kDa) in the myxobacterium *Stigmatella aurentica* (accession no. X87940; Bremaud *et al.*, 1997). With 927 aa (102.4 kDa), IF2 from *S. agalactiae* is among the larger IF2 proteins. A general feature observed when comparing IF2 sequences from different bacteria is a conserved C terminus and an N terminus that is highly variable in length and amino acid sequence. The amino acid sequences of IF2 from *S. agalactiae* and four closely related species were aligned (Fig. 2). A strong homology of the central and C-terminal part is observed, whereas the N-terminal part is more variable except for a homologous stretch including the first 50 aa. A segment, Ile¹⁸⁵–Glu²³⁷ in *S. agalactiae* IF2, was also found to be conserved between *S. agalactiae* and *S. pyogenes* IF2.

A structural model has been proposed for IF2 from *E. coli*, dividing it into six structural domains (Mortensen *et al.*, 1998). No function has been assigned to domain I (aa 1–157); domain II (aa 158–289) interacts with the

Fig. 5. Association between phylogenetic lineages and specific alleles of the central variable part of *infB* among 58 different clonal types of *S. agalactiae*. The dendrogram is adapted from Hauge *et al.* (1996) and was originally constructed on the basis

ribosomal 30S and 50S subunits (Moreno *et al.*, 1998, 1999); domain III (aa 290–389) is the link between domain II and IV; domain IV (aa 390–559) is responsible for the binding of GTP and exhibits GTPase activity together with the 50S ribosomal subunit; domain V (aa 560–671) interacts with the T arm of fMet-tRNA^{Met} (Yusupova *et al.*, 1996) and may, together with IF1, mimic the structure of elongation factor G at the ribosomal A site (Brock *et al.*, 1998; Åvarsson *et al.*, 1994); domain VI (aa 671–890) is believed to contain the fMet-tRNA^{Met} recognition site of IF2 (Gualerzi *et al.*, 1991). Domain III and/or V may interact with the initiation factor IF1 during translation initiation (Moreno *et al.*, 1999). Based on the strong homology of the central and C-terminal part, the six-domain model of *E. coli* IF2 can be adapted to IF2 from other bacteria regarding domains IV, V and VI, while the N-terminal domains can not be assigned due to lack of homology. Consequently, *S. agalactiae* IF2 can be divided into an N-terminal region (aa 1–427) and three domains, Glu⁴²⁸–Ala⁵⁹⁸, Asp⁵⁹⁹–Arg⁷¹⁰ and Val⁷¹¹–Lys⁹²⁷, being homologous to domains IV, V and VI, respectively, in *E. coli* IF2 (Fig. 2). The GTP-binding motif in domain IV of IF2 is well characterized and five regions, G1 to G5, essential for the activity have been described (Bourne *et al.*, 1991). As indicated in Fig. 2 the *S. agalactiae* IF2 contains the consensus sequence for each of the five regions.

The sequence of *infB* encoding the GTP-binding domain of IF2, which is homologous among species, is expected to be useful as a phylogenetic marker since *infB* is present in all bacteria as a single copy gene and has been shown to contain sufficient variation (Hedegaard *et al.*, 1999; Steffensen *et al.*, 1997). The DNA sequences of *infB* from Gram-positive species related to *S. agalactiae* and from *Mycoplasma* species supposed to originate from Gram-positive bacteria and, as outgroup, the Gram-negative *E. coli*, were aligned and the central part of the gene was used for phylogenetic analyses. A distance tree based on 16S rRNA sequences from the same species was constructed and compared to the topology obtained from *infB* sequence data (Fig. 3). The topology of the 16S rRNA tree obtained resembles the topology of 16S rRNA trees from the literature except for a difference in the position of *Enterococcus*. This is presumably because different parts of the 16S rRNA sequences and different methods for constructing the trees have been used. The genus *Enterococcus* groups with *Bacillus*, as in the work of Stackebrandt & Teuber (1988), whereas it groups with *Streptococcus* and *Lactococcus* in the work of Schleifer & Kilpper-Bälz (1987) and Olsen *et al.* (1994). The overall topology of the *infB* tree is equivalent to the 16S rRNA tree but supports the assumption that *Enterococcus* is closely related to the two other genera of Streptococci, *Strep-*

of 23 different characteristics. The serotype of the strains is indicated. The dagger symbol indicates that the full length sequence of *infB* was determined in these strains.

tococcus and *Lactococcus*, rather than to *Bacillus*. Consequently, our results indicate that the partial sequence of *infB* is useful as a molecular marker for phylogenetic studies within the Gram-positive bacteria.

Diversity in the sequence of *infB* among *S. agalactiae* strains

The *infB* sequence obtained for *S. agalactiae* strain 3076 was used to design oligonucleotide primers for amplification by PCR and sequencing of *infB* in an additional seven human isolates of *S. agalactiae*. The results divide the eight isolates into four alleles of *infB* sequences termed A to D (Fig. 4). Within *infB*, no deletions or insertions were observed and 14 polymorphic nucleotide positions were found, leading to three polymorphic amino acid positions in IF2, each with two possible amino acids. Residue 94 is Pro in allele A, B and C but Ser in allele D; residue 169 is His in allele A and B, and Tyr in allele C and D; and residue 744 is Ala in allele A, C and D, and Val in allele B (Fig. 4). The number of polymorphic nucleotide positions is lower than previously found for *infB* in *E. coli*, where a total of 66 polymorphic positions were detected in the sequence of *infB* among 10 human isolates (Steffensen *et al.*, 1997). The polymorphic nucleotide positions were unevenly distributed within *infB*, as the majority of the positions were located in the part encoding the GTP-binding domain and the C-terminal region of IF2. This resembles the distribution of variations found in *infB* from *E. coli*. In contrast, the observed number of polymorphic amino acid positions was higher than that reported among *E. coli*, where only a single amino acid variation was found (Steffensen *et al.*, 1997). The sequence results confirm our previous finding in *E. coli* that the sequence of *infB*/IF2 is extremely conserved within species.

To further assay for variation in the *infB* gene among *S. agalactiae* strains, we included an additional 50 strains for sequencing of the central variable region of *infB* encoding Lys³⁶¹–Gln⁵¹⁷. To obtain a diverse collection of *S. agalactiae*, the strains, together with the eight strains described above, were selected to represent each of the 58 clonal types described by Hauge *et al.* (1996). A total of six alleles were detected for the region of *infB* analysed. In addition to the four alleles shown in Fig. 4, which we will refer to as A to D as indicated, one allele termed E, represented by two closely related strains of clonal types 55 and 56, was identical to allele A except for a silent-site mutation of an A instead of T at position 1104, and one allele termed F, represented by a single strain of clonal type 28, was identical to allele A except for a T instead of G at position 1170 resulting in a conservative substitution of Glu³⁹⁰ to Asp. We observed a correlation between alleles of the region of *infB* analysed and the division into six major lineages, termed I through VI, based on our previous clonal typing of *S. agalactiae* (Hauge *et al.*, 1996) (Fig. 5). Thus, all 16 strains representing lineage V had the *infB* allele C, a cluster of six closely related strains within lineage II had allele B, whereas the remaining two distantly related strains of lineage II had *infB* allele A like the 11 strains

of lineage I, the two strains of lineage IV, and 16 out of 17 strains of lineage III analysed. The exceptional strain of lineage III had the unique *infB* allele F as described above. Two of four strains of lineage VI had *infB* allele E, strain 3098 belonging to clonal type 58 of this lineage had the unique allele D, and the strain included of clonal type 57, lineage VI had *infB* allele A. This association between certain *infB* alleles and distinct clonal lineages strongly supports our previous suggestion of dividing this population of *S. agalactiae* into distinct evolutionary lineages. Notably, division of serotype III strains into two genetically distinct subpopulations, lineage I and V, is confirmed by differences in their *infB* alleles. The fact that some lineages shared the same *infB* allele may indicate that these are more closely related than suggested by the topology of the dendrogram.

In conclusion, our results, based on variation in a partial sequence of *infB*, support the previously proposed clonal population structure of *S. agalactiae*. Though less discriminative, the central variable segment of *infB* proved to be useful as a molecular marker for phylogenetic studies of the species. The prokaryotic IF2 has been studied in several bacteria and the C-terminal part is found to be conserved between different species while the N-terminal part is characterized by its variability in both length and composition. Several different steps in the protein-synthesis mechanism can be affected by antimicrobial drugs and the species-specific N terminus of IF2 may be useful as a target for species-specific selective drugs.

ACKNOWLEDGEMENTS

The authors are grateful to Jens Jacob Hansen for providing the probe used for Southern blotting and to Katja Adolf for providing the sequence of *infB* from *Lactococcus lactis*. The sequence of *infB* from *Streptococcus pyogenes* was provided by B. A. Roe, S. P. Linn, L. Song, X. Yuan, S. Clifton, R. E. McLaughlin, M. McShan and J. Ferretti from the Streptococcal Genome Sequencing Project at the University of Oklahoma funded by USPHS/NIH grant AI38406.

This work was funded by grants from the Familien Hede Nielsens Fund and the Biotechnology Programme of the Danish Natural Science Research Council (28807-9502036, 9602401) to H. U. Sperling-Petersen and by grants from the Danish Medical Research Council (9702265) to M. Kilian.

REFERENCES

- Evarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., al-Karadaghi, S., Svensson, L. A. & Liljas, A. (1994). Three-dimensional structure of the ribosomal translocase: elongation factor G from *Thermus thermophilus*. *EMBO J* 13, 3669–3677.
- Ash, C., Farrow, J. A. E., Wallbanks, S. & Collins, M. D. (1991). Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small subunit rRNA sequences. *Lett Appl Microbiol* 13, 202–206.
- Bourne, H. R., Sanders, D. A. & McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature* 349, 117–127.
- Bremaud, L., Laalami, S., Derijard, B. & Cenatiempo, Y. (1997).

- Translation initiation factor IF2 of the myxobacterium *Stigmatella aurantiaca*: presence of a single species with an unusual N-terminal sequence. *J Bacteriol* **179**, 2348–2355.
- Brock, S., Szkaradkiewicz, K. & Sprinzl, M. (1998). Initiation factors of protein biosynthesis in bacteria and their structural relationship to elongation and termination factors. *Mol Microbiol* **29**, 409–417.
- Brombach, M., Gualerzi, C. O., Nakamura, Y. & Pon, C. L. (1986). Molecular cloning and sequence of the *Bacillus stearothermophilus* translational initiation factor IF2 gene. *Mol Gen Genet* **205**, 97–102.
- Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. (1978). Complete nucleotide sequence of a 16S rRNA gene from *Escherichia coli*. *Proc Natl Acad Sci USA* **75**, 4801–4805.
- Chung, C. T., Niemela, S. L. & Miller, R. H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci USA* **86**, 2172–2175.
- Cleat, P. H. & Timmis, K. N. (1987). Cloning and expression in *Escherichia coli* of the Ibc protein genes of group B streptococci: binding of human immunoglobulin A to the beta antigen. *Infect Immun* **55**, 1151–1155.
- Elish, M. E., Pierce, J. R. & Earhart, C. F. (1988). Biochemical analysis of spontaneous *fepA* mutants of *Escherichia coli*. *J Gen Microbiol* **134**, 1355–1364.
- Fraser, C. M., Gocayne, J. D., White, O. & 25 other authors (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.
- Friedrich, K., Brombach, M. & Pon, C. L. (1988). Identification, cloning and sequence of the *Streptococcus faecium* *infB* (translational initiation factor IF2) gene. *Mol Gen Genet* **214**, 595–600.
- Gold, L. (1988). Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Annu Rev Biochem* **57**, 199–233.
- Granlund, M., Oberg, L., Sellin, M. & Norgren, M. (1998). Identification of a novel insertion element, IS1548, in group B streptococci, predominantly in strains causing endocarditis. *J Infect Dis* **177**, 967–976.
- Green, C. J., Stewart, G. C., Hollis, M. A., Vold, B. S. & Bott, K. F. (1985). Nucleotide sequence of the *Bacillus subtilis* rRNA operon, *rrnB*. *Gene* **37**, 261–266.
- Gualerzi, C. O. & Pon, C. L. (1990). Initiation of mRNA translation in prokaryotes. *Biochemistry* **29**, 5881–5889.
- Gualerzi, C. O., Severini, M., Spurio, R., La Teana, A. & Pon, C. L. (1991). Molecular dissection of translation initiation factor IF2: evidence for two structural and functional domains. *J Biol Chem* **266**, 16356–16362.
- Hauge, M., Jespersgaard, C., Poulsen, K. & Kilian, M. (1996). Population structure of *Streptococcus agalactiae* reveals an association between specific evolutionary lineages and putative virulence factors but not disease. *Infect Immun* **64**, 919–925.
- Hedegaard, J., Steffensen, S. A. de A., Nørskov-Lauritsen, N., Mortensen, K. K. & Sperling-Petersen, H. U. (1999). Identification of *Enterobacteriaceae* by partial sequencing of the gene encoding translation initiation factor 2. *Int J Syst Bacteriol* **49**, 1531–1538.
- Hershey, J. W. B. (1987). Cellular and Molecular Biology. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, pp. 1386–1409. Edited by F. C. Neidhardt and others. Washington DC: American Society for Microbiology.
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkil, E., Li, B. C. & Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* **24**, 4420–4449.
- Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci USA* **86**, 8247–8251.
- Hubert, M., Nyengaard, N. R., Shazand, K., Mortensen, K. K., Lassen, S. F., Grunberg-Manago, M. & Sperling-Petersen, H. U. (1992). Tandem translation of *Bacillus subtilis* initiation factor IF2 in *E. coli*: overexpression of *infB*_{B,su} in *E. coli* and purification of alpha- and beta-forms of IF2_{B,su}. *FEBS Lett* **312**, 132–138.
- International Union of Biochemistry and Molecular Biology (1996). Prokaryotic and eukaryotic translation factors: ad hoc nomenclature subcommittee report. *Biochimie* **78**, 1119–1122.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Li, J., Kasper, D. L., Ausubel, F. M., Rosner, B. & Michel, J. L. (1997). Inactivation of the alpha C protein antigen gene, *bca*, by a novel shuttle/suicide vector results in attenuation of virulence and immunity in group B streptococcus. *Proc Natl Acad Sci USA* **94**, 13251–13256.
- Loenen, W. A. & Brammar, W. J. (1980). A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. *Gene* **10**, 249–259.
- Maitra, U., Stringer, E. A. & Chaudhuri, A. (1982). Initiation factors in protein biosynthesis. *Annu Rev Biochem* **51**, 869–900.
- Mikkelsen, B. M., Clark, M. E., Christiansen, G., Klintebaek, O. M., Nielsen, J. T., Thomsen, K. K. & Hjorth, J. P. (1985). The structure of two distinct pancreatic amylase genes in mouse strain YBR. *Biochem Genet* **23**, 511–524.
- Moreno, J. M., Kildsgaard, J., Siwanowicz, I., Mortensen, K. K. & Sperling-Petersen, H. U. (1998). Binding of *Escherichia coli* initiation factor IF2 to 30S ribosomal subunits: a functional role for the N terminus of the factor. *Biochem Biophys Res Commun* **252**, 465–471.
- Moreno, J. M., Dyrskjotersen, L., Kristensen, J. E., Mortensen, K. K. & Sperling-Petersen, H. U. (1999). Characterization of the domains of *E. coli* initiation factor IF2 responsible for recognition of the ribosome. *FEBS Lett* **455**, 130–134.
- Mortensen, K. K., Kildsgaard, J., Moreno, J. M., Steffensen, S. A., Egebjerg, J. & Sperling-Petersen, H. U. (1998). A six-domain structural model for *Escherichia coli* translation initiation factor IF2: characterization of twelve surface epitopes. *Biochem Mol Biol Int* **46**, 1027–1041.
- Musser, J. M., Mattingly, S. J., Quentin, R., Goudeau, A. & Selander, R. K. (1989). Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B streptococcus) causing invasive neonatal disease. *Proc Natl Acad Sci USA* **86**, 4731–4735.
- Nyengaard, N. R., Mortensen, K. K., Lassen, S. F., Hershey, J. W. B. & Sperling-Petersen, H. U. (1991). Tandem translation of *E. coli* initiation factor IF2 beta: purification and characterization *in vitro* of two active forms. *Biochem Biophys Res Commun* **181**, 1572–1579.
- Olsen, G. J., Woese, C. R. & Overbeek, R. (1994). The winds of (evolutionary) change: breathing new life into microbiology. *J Bacteriol* **176**, 1–6.
- Poulsen, K., Reinholdt, J. & Kilian, M. (1996). Characterization of the *Streptococcus pneumoniae* immunoglobulin A1 protease gene (*iga*) and its translation product. *Infect Immun* **64**, 3957–3966.
- Pritchard, D. G. & Lin, B. (1993). Group B streptococcal neuraminidase is actually a hyaluronidase. *Infect Immun* **61**, 3234–3239.
- Quentin, R., Huet, H., Wang, F. S., Geslin, P., Goudeau, A. & Selander, R. K. (1995). Characterization of *Streptococcus agalac-*

tiae strains by multilocus enzyme genotype and serotype: identification of multiple virulent clone families that cause invasive neonatal disease. *J Clin Microbiol* **33**, 2576–2581.

Remaut, E., Tsao, H. & Fiers, W. (1983). Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication. *Gene* **22**, 103–113.

Sacerdot, C., Dessen, P., Hershey, J. W., Plumbridge, J. A. & Grunberg-Manago, M. (1984). Sequence of the initiation factor IF2 gene: unusual protein features and homologies with elongation factors. *Proc Natl Acad Sci USA* **81**, 7787–7791.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schuchat, A. (1998). Epidemiology of group B streptococcal disease in the United States: shifting paradigms. *Clin Microbiol Rev* **11**, 497–513.

Schuchat, A. (1999). Group B streptococcus. *Lancet* **353**, 51–56.

Schleifer, K. H. & Kilpper-Bälz, R. (1987). Molecular and chemotaxonomic approaches to the classification of streptococci, enterococci and lactococci: a review. *Syst Appl Microbiol* **10**, 1–19.

Shazand, K., Tucker, J., Chiang, R., Stansmore, K., Sperling-Petersen, H. U., Grunberg-Manago, M., Rabinowitz, J. C. & Leighton, T. (1990). Isolation and molecular genetic characterization of the *Bacillus subtilis* gene (*infB*) encoding protein synthesis initiation factor 2. *J Bacteriol* **172**, 2675–2687.

Stackebrandt, E. & Teuber, M. (1988). Molecular taxonomy and

phylogenetic position of lactic acid bacteria. *Biochimie* **70**, 317–324.

Stålhammar-Carlemalm, M., Stenberg, L. & Lindahl, G. (1993). Protein rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. *J Exp Med* **177**, 1593–1603.

Steffensen, S. A. de A., Poulsen, A. B., Mortensen, K. K. & Sperling-Petersen, H. U. (1997). *E. coli* translation initiation factor IF2 – an extremely conserved protein: comparative sequence analysis of the *infB* gene in clinical isolates of *E. coli*. *FEBS Lett* **419**, 281–284.

Weisburg, W. G., Tully, J. G., Rose, D. L. & 9 other authors (1989). A phylogenetic analysis of the mycoplasmas: basis for their classification. *J Bacteriol* **171**, 6455–6467.

Wexler, D. E., Chenoweth, D. E. & Cleary, P. P. (1985). Mechanism of action of the group A streptococcal C5a inactivator. *Proc Natl Acad Sci USA* **82**, 8144–8148.

Wood, W. B. (1966). Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J Mol Biol* **16**, 118–133.

Yusupova, G., Reinbolt, J., Wakao, H., Laalami, S., Grunberg-Manago, M., Romby, P., Ehresmann, B. & Ehresmann, C. (1996). Topography of the *Escherichia coli* initiation factor 2/fMet-tRNA(f)(Met) complex as studied by cross-linking. *Biochemistry* **35**, 2978–2984.

Received 24 December 1999; revised 30 March 2000; accepted 10 April 2000.